MANAS Journal of Engineering
ISSN 1694-7398 | e-ISSN 1694-7398

*Volume 8 (Issue 1) (2020) Pages 55-58*

# Analysis of Turkish 6/49 lottery results

Alper Vahaplar[1], Murat Erşen Berberler[2,*]

[1]Dokuz Eylül University, Faculty of Science, Department of Statistics, İzmir, Türkiye, alper.vahaplar@deu.edu.tr, ORCID: 0000-0002-3566-3870

[2]Dokuz Eylül University, Faculty of Science, Department of Computer Science, İzmir, Türkiye, murat.berberler@deu.edu.tr, ORCID: 0000-0002-9227-2040

**ABSTRACT**

Turkish 6/49 Lottery is a chance game based on the selection of 6 of the 49 numbers by the draw machine, which has been of interest since the first day; such that statistics of this game, winning strategies, formulas etc., a number of popular books and newspaper articles have been published and some of the internet sites are now providing updated information to interested people; but a large part of these publications seem to be far from scientific point of view. In this study, all the draws in the past will be analyzed and the probability of being correctly guessed will be discussed by adhering to the theory of probability and statistical confidence tests.

**ARTICLE INFO**

***Research article***

*Received: 31.03.2020*
*Accepted: 22.05.2020*

***Keywords:***
*6/49 lottery,*
*probability,*
*confidence tests,*
*randomness*

**Corresponding author*

## 1. Introduction

Turkish 6/49 Lottery or officially "Sayısal Loto 6/49" has been played since 1996, November 6th and as of the date this paper is written, the number of draws reached 1107. The winners of this lottery are evident on Saturday evenings, 49 balls with numbers from 1 to 49 printed on them, 6 of them are drawn with the help of the lottery machine in the presence of a notary. Among the 6 numbers determined as the draw result, the ones who know 6, 5, 4, or 3 correct win lottery prize. There are no restrictions on the order in which the 6 numbers can be left out of the machine, the numbers from the drawing result are announced in ascending order, and therefore the computation is expressed in terms of mathematical combination. Considering that 6 numbers out of 49 numbers can be selected as 13.983.816 different types, the probability of correctly estimating the complete of 6 numbers that give big prize is approximately 1 in 14 million. The probability of guessing exactly 5 is obtained as $\frac{\binom{6}{5}\binom{43}{1}}{\binom{49}{6}} = \frac{258}{13983816}$, except that the case of 6 numbers are all correctly estimated. In a similar way, the likelihood of knowing exactly 4 with the exception of the cases of 6 and 5 are all correctly estimated is $\frac{\binom{6}{4}\binom{43}{2}}{\binom{49}{6}} = \frac{13545}{13983816}$, the probability of knowing exactly 3 is obtained by excluding the cases of 6, 5, and 4 are all correctly estimated is $\frac{\binom{6}{3}\binom{43}{3}}{\binom{49}{6}} = \frac{246820}{13983816}$. In accordance with all this information, in the following section, the weekly draws of Turkish 6/49 Lottery will be analyzed by statistical methods and the results will be interpreted.

## 2. Method

The $\chi^2$ goodness of fit test is one of the standard statistical tests used to verify whether two sets of data contain samples drawn from the same probability distribution. The chi-square statistic is a measure of how much the observed cell counts in a two-way table diverge from the expected cell counts [6]. This non-parametric test can be used even to test the hypothesis of no association between two or more groups or criteria and to test how likely the observed distribution of the data fits with the distribution expected [5]. The assumptions of $\chi^2$ goodness of fit test are listed as:

- The data must be randomly drawn from the population,
- There must be at least 5 expected frequencies in each group of categorical variable.
- The variables under consideration must be mutually exclusive.

In order to test the equiprobability of N individual numbers (which are 1 – 49 for Turkish 6/49 Lottery), frequencies are obtained. The expected probability for each can be considered as 1/N assuming the probability distribution is uniform. The goodness of fit test calculates a test statistic using observed frequencies and expected frequencies. For *n* draws of *k/N* type lottery, the expected count for any number *i* will be $E_i = \frac{nk}{N}$ [3].

$$\chi^2 = \sum_{i=1}^{N} \frac{(O_i - E_i)^2}{E_i} \qquad (1)$$

The null hypothesis of $\chi^2$ is that observed values are equal to the expected values:
$H_0 : O_i = E_i$ (Observed and Expected data are the samples of the same distribution)

$H_a : O_i \neq E_i$

The test outputs $\chi 2$ value and a p-value to compare to the desired significance level ($\alpha$). If $p<\alpha$, then H0 is rejected and accepted otherwise.

In this study, 1107 draws of Turkish 6/49 Lottery was concerned. For 1107 draws, the frequency table of 49 numbers and p-value of $\chi 2$ test is as follows:

**Table 1.** *Frequencies of 49 lotto numbers over 1107 draws*

| *i* | *Oᵢ* | *i* | *Oᵢ* | *i* | *Oᵢ* | *i* | *Oi* | *i* | *Oᵢ* |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 149 | 11 | 136 | 21 | 151 | *31* | *122* | 41 | 143 |
| 2 | 138 | 12 | 141 | 22 | 144 | *32* | *144* | 42 | 130 |
| 3 | 137 | 13 | 136 | 23 | 130 | *33* | *120* | 43 | 104 |
| 4 | 127 | 14 | 141 | 24 | 128 | *34* | *131* | 44 | 130 |
| 5 | 137 | 15 | 133 | 25 | 133 | *35* | *129* | 45 | 123 |
| 6 | 122 | 16 | 150 | 26 | 148 | *36* | *149* | 46 | 134 |
| 7 | 133 | 17 | 144 | 27 | 137 | *37* | *115* | 47 | 143 |
| 8 | 138 | 18 | 156 | 28 | 127 | *38* | *158* | 48 | 125 |
| 9 | 142 | 19 | 133 | 29 | 128 | *39* | *137* | 49 | 134 |
| 10 | 129 | 20 | 136 | 30 | 137 | *40* | *150* | | |

p-value = 0,809578063, $\chi^2$ = 39.322 and df = 48

If the level of significance ($\alpha$) is accepted as 0.05, the test result can be interpreted as there is no sufficient evidence to reject the claim in the null hypothesis. In other words, from this perspective, 1107 draws of Turkish 6/49 Lottery seem to be fair.

The same analysis is applied to packages of 25, 50 and 100 consecutive draws in total of 1100 draws. The overall is again proved to be fair but for some sequences of draws seemed suspicious in terms of p-values being less than $\alpha$=0.05.

**Table 2.** *p-values in short sequences of draws.*

| Draws | p-value | Draws | p-value |
|---|---|---|---|
| 351 - 375 | 0,010509424 | 401 - 500 | 0,938341051 |
| 351 - 400 | 0,044315743 | 501 - 600 | 0,805498036 |
| 651 - 700 | 0,00499506 | 601 - 700 | 0,07817169 |
| 1 - 100 | 0,360785414 | 701 - 800 | 0,86715685 |
| 101 - 200 | 0,448803486 | 801 - 900 | 0,871352253 |
| 201 - 300 | 0,528784303 | 901 - 1000 | 0,84498258 |
| 301 - 400 | 0,378982582 | 1001 - 1100 | 0,99491968 |

This situation can be interpreted as a result of randomness in small samples. As sample sizes are enlarged, p-values are determined to increase.

The total number of possible combinations of k numbers chosen from a set of N numbers is given by the combinatorial coefficient

$$\binom{N}{k} = \frac{N!}{k!\,(N-k)!} \qquad (2)$$

where N = 49 and k = 6 for Turkish 6/49 Lottery. The players get a reward in cases of truly guessing 3, 4, 5 or all of the 6 numbers drawn. The asymptotic distribution of $\chi^2$ statistics is neither uniform nor chi-square because the numbers are drawn *without replacement*. When one of 49 numbers is selected, there is no chance for the same number to be selected again in the same draw [3],[7].

Let X be the random variable defined as number of matches out of k randomly drawn numbers among N numbers with equiprobability. In this case, the distribution of X is said to be hypergeometric. The probability of matching *i* numbers of *k* numbers drawn in N numbers is [1],[4].

$$P[X = i] = \frac{\binom{k}{i}\binom{N-k}{k-i}}{\binom{N}{k}} \qquad (3)$$

Let Y be the random variable denoting the sorted outcome vector of a draw .Y = [$Y_{(1)}$, $Y_{(2)}$, …$Y_{(k)}$] where $Y_{(i)}$ is the random variable corresponding to the $i^{th}$ element in the sorted outcome. Recall that $Y_{(1)} < Y_{(2)} < \ldots < Y_{(k)}$.Under these assumptions, the probability of $i^{th}$ number in the sorted outcome vector having the value of *r* is derived from Equation (3) as:

$$P\left[Y_{(i)} = r\right] = \frac{\binom{r-1}{i-1}\binom{N-r}{k-i}}{\binom{N}{k}} \qquad (4)$$

It is clear that if $Y_{(i)} = r$, then only $i - 1$ numbers fall between 1 and $r - 1$, and $k - i$ numbers fall between $r + 1$ and N. Using Eq. (4) to determine the expected value of Y(i) in terms of E[x]=∑x.p(x) [1]:

$$E[Y_{(i)}] = \frac{\sum_{r=i}^{N-k+i} r\binom{r-1}{i-1}\binom{N-r}{k-i}}{\binom{N}{k}} = \frac{(N+1)i}{(k+1)} \tag{5}$$

So the expected values of Turkish 6/49 lottery are determined as

μ = [50/7, 100/7, 150/7, 200/7, 250/7, 300/7]

μ = [7.142857143, 14.28571429, 21.42857143, 28.57142857, 35.71428571, 42.85714286]

which are very close to the average value of 1107 draws

Y = [7.044263776, 14.08852755, 21.15898826, 28.35772358, 35.4263776, 42.64046974]

A t-test is applied to the expected values and the previous results using $H_0$ = the difference in means is equal to 0. The results are as follows:

t = 0.027775, df = 9.9999, p-value = 0.9784
alternative hypothesis: true difference in means is not e qual to 0
95 percent confidence interval: -16.94890 17.37678
sample estimates:
mean of x mean of y
25.00000 24.78606

Commenting on p-value result of the test (0.9874), Turkish 6/49 lottery is said to be fair in perspective of hypergeometric distributed random variables.

Another study on fairness or equiprobability on lottery games results belongs to Drakakis et. al. [2]. This study focuses on the minimal distance of consecutive result of a draw. Let $Y_{(1)}$, $Y_{(2)}$, …$Y_{(k)}$ be numbers within range 1…N. The minimal distance d is defined as

$$d = \min_{1\le i\le j\le k} |Y_{(j)} - Y_{(i)}| \tag{6}$$

As an example, if the sorted outcome vector Y has numbers (4, 17, 19, 23, 30, 44) drawn, then d = 19 – 17 = 2 for this individual sample. Assuming the numbers are randomly chosen and the outcome vector Y is uniformly distributed within the sample space of $\binom{N}{k}$ k-tuples, the probability distribution of random variable d is

$$P(d \le r) = 1 - \frac{\binom{N-(r-1)(k-1)}{k}}{\binom{N}{k}} \tag{7}$$

At this stage, the minimal distance frequencies of previous 1107 draws are computed and compared with the expected values using $\chi^2$ goodness of fit test.

**Table 3.** *Observed and expected values of distances and their probabilities*

| d | Observed | Expected | P(d<r) |
|---|---|---|---|
| 1 | 548 | 548,1847 | 0,000000 |
| 2 | 295 | 300,5365 | 0,495198 |
| 3 | 148 | 151,8123 | 0,271487 |
| 4 | 72 | 68,86258 | 0,137139 |
| 5 | 36 | 26,94897 | 0,062207 |
| 6 | 7 | 8,507166 | 0,024344 |
| 7 | 1 | 1,910123 | 0,007685 |
| 8 | 0 | 0,231077 | 0,001726 |

The p-value of $\chi^2$ goodness of fit test is calculated as 0.6658057 which gives no doubt about the randomness / fairness or equiprobability of Turkish 6/49 Lottery for the previous 1107 draws.

## 3. Summary and conclusion

Based on the literature Turkish 6/49 Lottery Game results are investigated in terms of fairness and randomness using different methods. The previous data of 1107 games of lottery has been analyzed and inspected to be fair. In the first part of the study basic $\chi^2$ goodness of fit test is introduced and the application on Turkish data set is presented. For the overall analysis of the lottery results gave no evidence on being unfairness as of test result. But in particular investigation of consecutive weeks, small p-values like 0.00499 which constitutes strong statistical evidence to conclude that the observed data are statistically different from the expected values. As a further research, those specific results in the determined weeks can be analyzed in different tests or methods.

As an alternative research of randomness, the sorted outcomes of the results for each week are considered to be a random variable in hypergeometric distribution. The expected values and the observed results are tested. According to the results, $H_0$ is failed to reject, meaning no conclusions can be made on "observed values are statistically different from the expected values".

Lastly, the minimum distance metric in the 6-tuples of outcomes were used for testing the empirical and observed values. Applying $\chi^2$ goodness of fit test again resulted as there is no evidence for rejecting the null hypothesis that both data samples are drawn from the same distribution.

The statistical computations and hypothesis tests are performed in R Version 3.4.3 using R-Studio Version 1.1.419. (https://www.R-project.org/)

## 4. Further research

Using the historical data of Turkish 6/49 lottery, some predictions can be made for the future Sample size of 1107 may be insufficient to conclude on some hypothesis. So the test explained above can be repeated for large samples or on other similar game results.

As the sample size increases, some prediction algorithms can be studied to guess the numbers in 6-tuples.

## References

[1] Coronel-Brizio H.F., Hernández-Montoya A.R., Rapallo F., Scalas E., "Statistical auditing and randomness test of lotto k/N-type games", Physica A: Statistical Mechanics and its Applications, 387(25), (2008), 6385-6390.

[2] Drakakis K., "A note on the appearance of consecutive numbers amongst the set of winning numbers in Lottery", Facta Universitatis: Mathematics and Informatics, 22(1), (2007), 1–10.

[3] Genest C., Lockhart R.A. and Stephens, M.A., "Chi-square and the lottery", Journal of the Royal Statistical Society, Series D, 51, (2002), 243—257.

[4] Neufeld E., Goodwin S., "The 6-49 lottery paradox", Computational Intelligence, 14(3), (1998), 273-286.

[5] Rana R, Singhal R., "Chi-square test and its application in hypothesis testing", Journal of the Practice of Cardiovascular Sciences, 1, (2015), 69-71.

[6] Moore, David S., George P. McCabe, and Bruce A. Craig. "Introduction to the Practice of Statistics", New York: W.H. Freeman, (2009).

[7] Harry Joe, "Tests of uniformity for sets of lotto numbers", Statistics & Probability Letters, 16(3), (1993), 181-188.